Video-Based Quantification of Gait Impairments in Parkinson's Disease Using Skeleton-Silhouette Fusion Convolution Network

Qingyi Zeng, Peipei Liu, Ningbo Yu, *Member, IEEE*, Jialing Wu, Weiguang Huo, and Jianda Han

***Abstract*— Gait impairments are among the most common hallmarks of Parkinson's disease (PD), usually appearing in the early stage and becoming a major cause of disability with disease progression. Accurate assessment of gait features is critical to personalized rehabilitation for patients with PD, yet difficult to be routinely carried out as clinical diagnosis using rating scales relies heavily on clinical experience. Moreover, the popular rating scales cannot ensure fine quantification of gait impairments for patients with mild symptoms. Developing quantitative assessment methods that can be used in natural and home-based environments is highly demanded. In this study, we address the challenges by developing an automated video-based Parkinsonian gait assessment method using a novel skeleton-silhouette fusion convolution network. In addition, seven network-derived supplementary features, including critical aspects of gait impairment (gait velocity, arm swing, etc.), are extracted to provide continuous measures enhancing low-resolution clinical rating scales. Evaluation experiments were conducted on a dataset collected with 54 patients with early PD and 26 healthy controls. The results show that the proposed method accurately predicted the patients' unified Parkinson's disease rating scale (UPDRS) gait scores (71.25% match on clinical assessment) and discriminated between PD patients and healthy subjects with a sensitivity of 92.6%. Moreover, three proposed supplementary features (i.e., arm swing amplitude, gait velocity, and neck forward bending angle) turned out to be effective gait dysfunction indicators with Spearman correlation coefficients of 0.78, 0.73, and 0.43 matching the rating scores, respectively. Since the proposed system requires only two smartphones, it holds a significant benefit for home-based quantitative assessment of PD, especially for detecting early-stage PD. Furthermore, the proposed supplementary features can enable high-resolution assessments of PD for providing subject-specific accurate treatments.**

***Index Terms*— Parkinson's disease, gait impairments, video-based assessment, spatial-temporal graph convolutional network.**

Manuscript received 21 September 2022; revised 7 March 2023, 22 May 2023, and 7 June 2023; accepted 16 June 2023. Date of publication 7 July 2023; date of current version 12 July 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB4700203, in part by the National Natural Science Foundation of China under Grant U1913208, in part by the Tianjin Key Medical Discipline (Specialty) Construction Project under Grant TJYXZDXK-052B, in part by the Tianjin Health Research Project under Grant TJWJ2022MS031, and in part by the Fundamental Research Funds for the Central Universities. *(Qingyi Zeng, Peipei Liu, and Ningbo Yu contributed equally to this work.) (Corresponding authors: Jialing Wu; Weiguang Huo; Jianda Han.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Tianjin Huanhu Hospital, Tianjin, China, under Approval No. ChiCTR1900025372.

Qingyi Zeng and Weiguang Huo are with the College of Artificial Intelligence, Nankai University, Tianjin 300350, China (e-mail: zengqingyi@mail.nankai.edu.cn; weiguang.huo@nankai.edu.cn).

Peipei Liu and Jialing Wu are with the Department of Neurology, Tianjin Huanhu Hospital, Tianjin 300350, China, and also with the Tianjin Key Laboratory of Cerebral Vascular and Neurodegenerative Diseases, Tianjin Neurosurgical Institute, Tianjin 300350, China (e-mail: liupeipei2008@sina.com; wywjl2009@hotmail.com).

Ningbo Yu and Jianda Han are with the College of Artificial Intelligence, Nankai University, Tianjin 300350, China, and also with the Institute of Intelligence Technology and Robotics Systems, Shenzhen Research Institute of Nankai University, Shenzhen 518083, China (e-mail: nyu@nankai.edu.cn; hanjianda@nankai.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3291359, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3291359

I. Introduction

PARKINSON'S disease (PD) is the second most common neurodegenerative disorder affecting over 10 million individuals worldwide [1], [2], [3]. The main symptoms of PD, including bradykinesia, rigidity, tremor, and postural instability, usually cause gait impairments for patients [3]. Since gait impairments, such as reduced arms swing, step length, and gait speed, appear in the earliest stage of PD, they are considered important biomarkers for discriminating early PD. Moreover, gait impairments worsen with disease progression and markedly impact the mobility and quality of life of patients with PD. Accurate assessment of gait impairments is crucial for initiating neuroprotective therapies in the earliest stage [4] and tailoring subject-specific treatments during the whole disease stage. Traditional clinical assessments of PD are mainly based on rating scales such as the unified Parkinson's disease rating scale (UPDRS) [5] and Hoehn and Yahr (H&Y) [6], relying on clinical experience and resources and subject to inter- and intra-rater subjectivity across examiners. Moreover, rating scales cannot ensure high-resolution assessments to

reflect changes in gait impairments for patients with early PD. Hence, developing automated and quantitative PD assessment methods that can be easily used in natural and home-based environments is of great importance.

To date, many sensor-based methods have been proposed for quantitative assessments of PD gait impairments. Researchers utilized sensor information such as ground reaction force [7], [8], motion capture [9], [10], surface electromyography (EMG) [11], [12] to analyze gait features for the early diagnosis of PD. However, such rich and accurate data highly depends on laboratory-based experiments conducted in clinics with the help of trained experts, which brings difficulties for routine assessments.

For non-laboratory assessments, wearable sensors have been widely used due to advantages such as lightweight and ease of use [13], [14]. Accelerometer [15], inertial sensors [16], [17], [18], mechanomyography [18], [19] have been used for measurement of bradykinesia, rigidity, and tremor. These sensors were placed on the human body to monitor kinematic features or muscle activities. Since PD affects many parts of the body and causes a wide range of motion problems, an accurate assessment of PD severity relies on a comprehensive analysis of human motion features. Hence, such methods need an important number of wearable sensors for assessing gait impairments, which limits their applications in home-based and natural scenarios.

With the rapid development of computer vision, a series of vision-based methods have been proposed to quantify PD. In [20], [21], [22], and [23], 3D human poses were extracted using depth cameras (e.g., Microsoft Kinect) for PD assessment. Movement features, such as stride length, stride velocity and arm swing, were estimated based on human poses. Meanwhile, machine learning algorithms, such as support vector machine (SVM), random forest (RF), and K-nearest neighbours (KNN), were used to discriminate between PD patients and healthy subjects.

Recently, a small body of work has tried to use videos filmed using cameras and smartphones to diagnose PD. In [24] and [25], authors conducted early PD screening using videos filming subjects' gaits. In [26], [27], and [28], authors proposed three neural network-based methods for predicting MDS (Movement Disorder Society)-UPDRS scores using smartphone videos. The smartphone video-based methods hold significant potential for achieving remote home-based PD diagnosis due to advantages such as low cost, contactless, convenience, and flexibility. Meanwhile, there is still a large space to improve the prediction accuracy and robustness with respect to the diversity of human gaits.

There are two commonly used types of methods for analyzing PD motor disorders based on videos. The first one usually employs statistical analysis to distinguish PD patients [25], [29]. These methods have good interpretability. However, in many cases, it is difficult to discover explicit relationships/models between PD severity and motion features due to the diversity and variability of PD symptoms. For the second one, machine/deep learning approaches are used to estimate the MDS-UPDRS scores [26], [27], [28], [30]. These methods can provide relatively high classification accuracy. However, few works focused on further extracting quantitative features for the high-resolution assessment of PD symptoms.

In this paper, a skeleton-silhouette fusion neural network is proposed to automatically predict the MDS-UPDRS scores based on gait videos filmed using smartphones (see Fig.1). The proposed model consists of a skeleton stream and a silhouette stream. For the former, it takes pose graphs as inputs, which are constructed by the 2D human poses estimated from the gait videos, and utilizes a spatial-temporal graph convolutional network (ST-GCN) to extract the skeleton feature vectors. For the latter, it takes the Long-Term GEI (Gait Energy Image) as input, which consists of several silhouette images sequentially extracted from the gait videos, and uses a VGG network to extract silhouette feature vectors. Then, the two feature vectors are fused for predicting the MDS-UPDRS gait scores. Moreover, saliency values are derived from the proposed neural network to understand which body parts contribute more to the ensuing correct prediction of the gait scores. Finally, seven supplementary features are extracted to validate the effectiveness of the saliency values. The proposed method was evaluated on a dataset collected with 54 patients with early PD and 26 healthy controls. The proposed model achieved an appreciable accuracy of 71.25%, which outperforms other state-of-the-art (SOTA) methods. The neck, torso and arm are the three body parts with the largest saliency, and the arms swing amplitude, the torso movement velocity and the neck forward bending angle were found to be significantly correlated to the gait scores with the Spearman correlation coefficient of 0.78, 0.73 and 0.43. The supplementary features provide credible quantification metrics with higher resolution.

The main contribution of our work can be summarized as follows:

- The skeleton-silhouette fusion convolution network is proposed to quantitatively assess early-stage Parkinsonian gaits using smartphone-based videos.
- The saliency values are derived from the proposed network for understanding the contribution of features related to each body part in correctly predicting the gait scores.
- Seven supplementary features are extracted to provide continuous measures of gait impairments to enhance low-resolution clinical rating scales.

II. Method

Fig. 1 illustrates the overall pipeline of the proposed skeleton-silhouette fusion convolution network based method. Participants are required to walk approximately 5m away in a clinical or home-based environment. Meanwhile, participants' gaits are recorded using two smartphones from both the sagittal view and the coronal view. Note that only the videos recorded from the sagittal view are used in the proposed network for the assessment of gait impairments. The videos recorded from the coronal view are used to ensure the left leg and right leg are correctly estimated by the OpenPose. Both human silhouettes and skeleton sequences are extracted from the recorded sagittal videos by Mask R-CNN [31] and OpenPose [32]. Then, the silhouette sequences are transformed

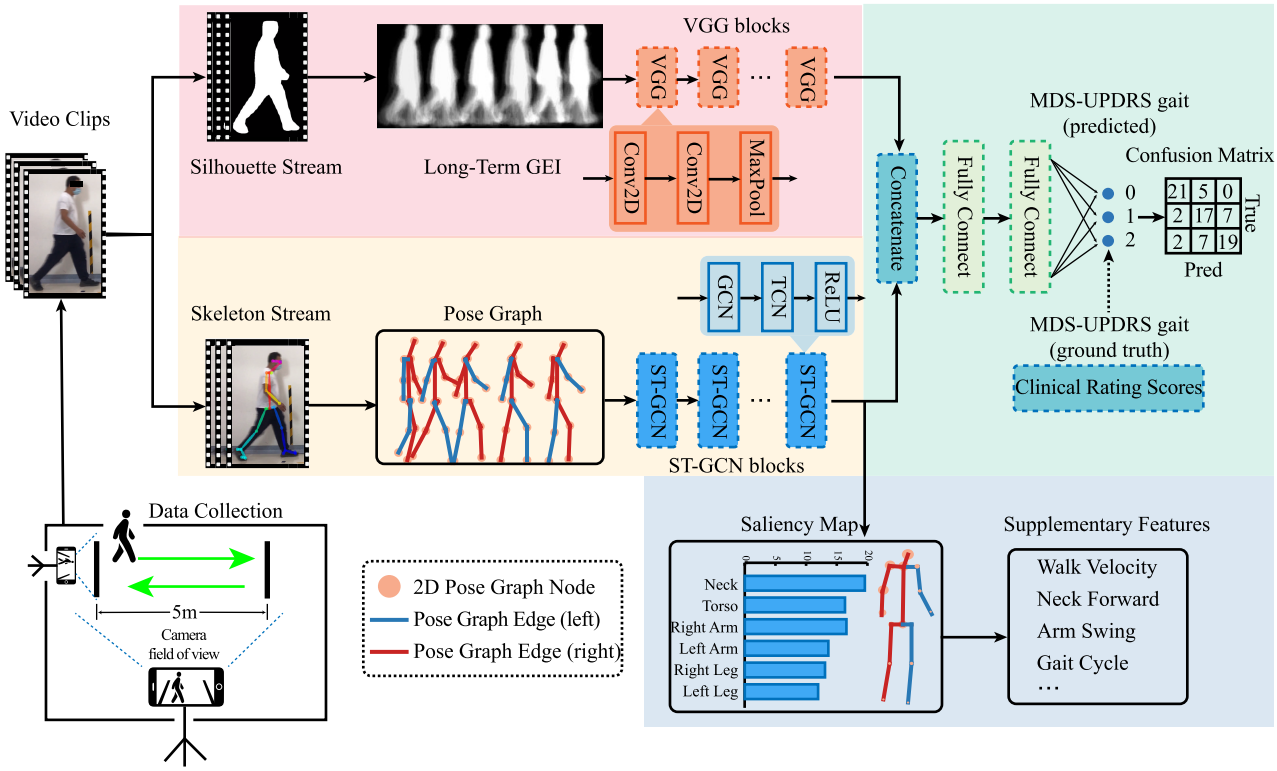


Fig. 1. The proposed framework. The silhouettes and skeletons are extracted from gait videos. Then the Long-Term GEI and pose graphs are constructed and put into the skeleton-silhouette fusion neural network to predict the MDS-UPDRS gait scores. Besides the prediction results, the framework can also give the saliency values of six body parts. Furthermore, supplementary features are extracted for fine-grained assessment of gait impairments.

into a form of GEI called Lone-Term GEI [33], while the skeleton sequences are arranged into undirected pose graphs. The two-stream network consists of two parts: a silhouette stream and a skeleton stream. The former detects gait cycle-related lower limb motion features by 2D convolution, while the latter extracts spatial and temporal Parkinsonian gait features using graph convolution. Then, all features are fused by fully connected layers to predict the MDS-UPDRS gait scores of the participants. Furthermore, to visualize which joints contribute more to the prediction task, the saliency values are derived from the skeleton stream. Finally, seven supplementary features are selected for verifying the saliency values.

A. Subjects and Experimental Protocols

54 patients with PD and 26 healthy subjects were recruited for this study. Table I shows their basic information. All gait scores are given by board-certified clinicians. The gaits of all healthy subjects are also assessed by board-certified clinicians to ensure that their gaits are normal, i.e., with a score of 0. Since this study mainly focuses on the assessment of early-stage Parkinsonian gaits, all recruited PD patients are with UPDRS scores of 1 or 2. Note that PD patients with scores of 3 and 4 cannot walk independently [5]. The data collection is carried out in Huanhu Hospital, Tianjin, China. All patients' data are collected during the off-medication state. All videos are recorded using two smartphones with a 720p resolution and 25fps frame rate. All procedures are approved by the Institutional Review Board of Tianjin Huanhu Hospital with the reference number ChiCTR1900025372. All participants gave their consent for the experimental procedure.

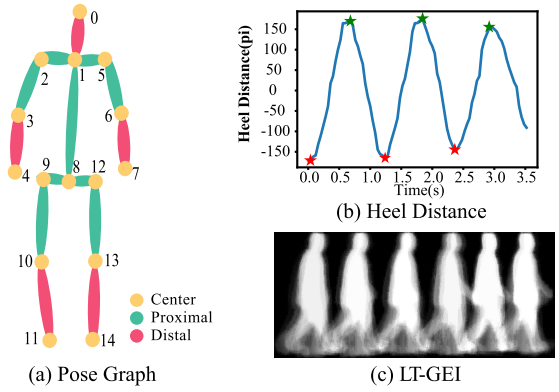
As shown in Fig. 1, participants are required to walk approximately 5m away and toward the examiner three times during one test. Each recorded video is divided into 6 equal clips in order to efficiently use the acquired data and facilitate subsequent processing. The turning parts are not included in the study. A total of 480 video clips are finally involved for training and validation.

B. Data Processing

1) Graph Construction: The graph is a kind of data that is suitable to describe human kinematics, as it can reflect not only the position of each joint but also the physical connections between joints. An undirected spatial-temporal graph can be expressed as $G = (V, E)$, where V and E denote the node and edge sets, respectively. Human joints extracted from gait videos are employed to construct the pose graphs. As shown in Fig. 2 (a), a 2D human pose graph, which consists of 15 joints, can be estimated from the videos filmed in the sagittal view using the OpenPose framework [32]. Due to the occlusion problem, the left and right sides of the lower limbs are incorrectly exchanged in some specific cases.

TABLE I
BASIC STATISTICS OF THE PARTICIPANTS

Score	All	0	1	2
Number of Participants	80	26	26	28
Number of Video Clips	480	156	156	168
Sex (M/F)	43/37	11/15	16/10	16/12
Age (years±SD)	61.4±6.2	61.5±5.1	59.7±6.7	63.1±6.2
Symptom Duration (years±SD)	4.48±3.04	-	4.11±3.27	4.82±2.70



(a) Pose Graph (b) Heel Distance (c) LT-GEI

Fig. 2. Gait data preprocessing. (a) Pose graph with proximal-distal configuration. (b) Distance between two heels (unit: pixel). The time interval between the red stars and green stars is recognized as gait cycles. (c) An example of the LT-GEI during 3 gait cycles.

To address this problem, the coronal videos are used to detect and correct the incorrectly estimated frames. Note that the eyes and ears are omitted because they are less related to the gait movements. Meanwhile, the heels and toes are also omitted to avoid excessive measurement noises [34].

2) Proximal-Distal Configuration: The node set V is composed of all 2D joint positions extracted from the gait videos. As for the edge E, a proximal-distal configuration is designed to divide the edges into three categories. Previous work [26] considered all natural connections of joints as same edges. Nevertheless, according to [35], the proximal-to-distal motion pattern was observed in walking and running, revealing the fact that different joints and muscle groups assume different functions in human movements. Given that, we propose to divide the joint connections into temporal, proximal, and distal categories to distinguish different joints and their connections. Note that joints in consecutive frames are connected by temporal edges to represent temporal dynamics. According to the definition of the anatomical position, take the center of the trunk as the origin, the proximal refers to joints and connections near the origin, and the distal refers to the joints and connections far from the origin. Hence edges within the torso, the upper arms, and the thighs parts are assigned to the proximal group, while the ones within the neck, forearms, and lower limbs parts are assigned to the distal group. As shown in Fig. 2 (a), the yellow, green, and red nodes represent the central, proximal, and distal categories, respectively.

3) Long-Term GEI: For the sake of avoiding extra noises, the heels and toes are omitted in the previous section. However, recent studies have shown that foot movements are important in the assessment of parkinsonian gaits [36], [37]. To replenish lower limb motion features in the model, human silhouettes are involved since silhouettes have been proven to be an alternative to motion capture due to their high accuracy [29]. The Mask R-CNN framework [31] is used for human silhouette segmentation from sagittal videos. Inspired by [33], the Long-Term GEI (LT-GEI) is proposed to extract foot movements within three sequential gait cycles. GEI is commonly used for gait analysis and has the ability to present both dynamic and static movement features in one gait cycle. The generation method of the proposed LT-GEI is expressed as follows:

$$GEI(x, y) = \frac{1}{N}\sum_{n=1}^{N} I_c(x, y, n) \tag{1}$$

where $I_c(x, y, n)$ represents the human silhouette of the n^{th} frame. N represents the number of frames contained in an image. The transitional GEI cannot represent the reduced stride length, a typical feature of Parkinsonian gait [38]. However, the proposed LT-GEI is intended to represent features within and between several gait cycles. The core idea of LT-GEI is to compose 6 GEIs extracted from 3 gait cycles into one image. The construction method is expressed as follows:

$$LT\text{-}GEI(x, y) = \sum_{i=0}^{M} \frac{1}{T_{i+1} - T_i} \sum_{n=T_i}^{T_{i+1}} I_c(x, y, n) \tag{2}$$

where $I_c(x, y, n)$ represents the segmentation result of the n^{th} frame of a gait video. T_i represents the frame in which the i^{th} heel contact event is detected. M represents the number of gait cycles required by an LT-GEI. The heels are identified from the silhouettes, and then the distance d_i is calculated as follows:

$$d_i = X_L - X_R \tag{3}$$

where X_L and X_R represent the horizontal positions of the left and right heel in frame i. As shown in Fig.2 (b), the red and green stars are identified as time stamps T_i of the occurrence of heel contact. Fig. 2 (c) is an example of the LT-GEI. By combining 6 GEIs into one image, the LT-GEIs can characterize gait cadence as well as gait variation.

C. Feature Extraction

The skeleton-stream fusion neural network is designed to extract features from the poses graphs and LT-GEIs and to predict the gait scores. An instantiation of the network is listed in Table II. After testing several two-stream structures with different numbers of ST-GCN blocks and VGG blocks, we selected the model structure, which has achieved a trade-off between the parameter size and the accuracy.

TABLE II
THE STRUCTURE OF THE SKELETON SILHOUETTE FUSION NETWORK

Stage	Skeleton Stream	Silhouette Stream	Output Sizes
raw clip	-	-	Skel: 2×64×15 Silh: 1×224×224
Conv	[15×9, 64]×2 [15×9, 128]×2 [15×9, 256]×2	[3^2, 64]×2 [3^2, 128]×2 [3^2, 256]×2 [3^2, 512]×3 [3^2, 512]×3	Skel: 256×8×15 Silh: 512×7×7
Pooling+FC Flatten+FC	[256,256]	[25088,4096] [4096,256]	Skel: 256×1 Silh: 256×1
Concat	-		Fusion: 512×1
FC	[512,3]		Predict: 3×1

For the skeleton stream (Skel), $[S \times T, C] \times n$ represents the spatial kernel, the temporal kernel, the number of channels, and the number of ST-CN blocks, respectively. For the silhouette stream (Silh), the VGG16 convolutional layers are employed. After the convolutional operation, the dimension of the feature extracted is reduced by pooling, flattening and an FC (fully connect) layer. Then they are fused by the concatenate layer (Concat) and two FC layers. The parameters and output dimensions of each stage are listed in the table. The ablation study was done to show how the model size was chosen (see attached materials).

1) Skeleton Stream: Inspired by the successful application of ST-GCN in action recognition [39], the ST-GCN model is investigated to extract spatial and temporal Parkinsonian gait features from the pose graph sequences. There are two types of basic operations in the ST-GCN model: spatial graph convolution and temporal convolution. The former is introduced to capture posture symptoms such as anticollis and camptocormia [40], while the latter is utilized to model time-dependent motion symptoms such as slow pace and arm swing asymmetry [10], [38].

Following the concepts of the graph convolutional network proposed in [41], the graph convolutional operation of the proposed ST-GCN is expressed as follows:

$$X_{out} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X_{in} W \tag{4}$$

where $X_{out} \in \mathbb{R}^{15*2*L}$ and $X_{in} \in \mathbb{R}^{15*2*L}$ are the output and input matrices of the graph convolution layer, respectively. L denotes the input sequence length. $A \in \mathbb{R}^{15*15}$ represents the normalized adjacency matrix of the pose graph. $D \in \mathbb{R}^{15*15}$ is the degree matrix, which is a diagonal matrix calculated by $D_{ii} = \Sigma_j A_{ij}$. W denotes a layer-specific learnable weight matrix. The temporal, proximal, and distal connections are treated differently by using undirected weighted graphs. The proximal connections are denoted as graph edges with a weight of 2, while the distal connections are represented by the edges with a weight of 3. For example, the connection between Joint 1 and Joint 2 is defined as the proximal connection, then we set $A(1,2) = A(2,1) = 2$. The temporal connection is implemented by 1D convolution. For a joint J_i, the temporal sequence can be represented by $J_{i1}, J_{i2}, J_{i3}, \ldots, J_{it}$, then the temporal convolution over this joint is implemented by 1D convolution using a 1×9 kernel. Following the concept of 1D convolution, the temporal convolution is expressed as follows:

$$Y = \sum_{k=0}^{N} W_k \otimes X_{out} + B \tag{5}$$

where Y denotes the output matrix of the temporal convolution layer. N is the number of convolution kernels, W_k shows the learnable kernel, B is the learnable bias.

2) Silhouette Stream: Previous studies have shown that patients with PD have higher step length variability compared to healthy people [38], [42]. To detect the gait cycle-related spatial features from the LT-GEI, the convolutional neural network (CNN) is employed. The VGG16 architecture is chosen to achieve a compromise between prediction accuracy and parameter scale. As shown in Tab II, the silhouette stream consists of 5 VGG blocks and 1 fully connected layer. The VGG network is good at image classification. However, due to the huge hyperparameter size, it is computationally expensive. In the VGG16 network, the convolution layers can extract patterns from the pixels, while the fully connect layers pick and fuse these patterns to adapt to specific classification domains. The model parameters pretrained on the ImageNet have proved to be effective in detecting natural patterns. Hence the 5 VGG blocks are initialized with these parameters to help detect natural body patterns such as neck, arms, and legs. During the model training procedure, only fully connect layers parameters are updated to detect Parkinsonian gait impairments. Thus the computation cost can be reduced.

D. Saliency Analysis and Supplementary Features

To better understand the features of which body parts play a more important role in predicting patients' UPDRS scores, the saliency value is defined as a quantification index. The saliency value of each joint is derived from the output of the ST-GCN blocks. Given a pose graph sequence input with shape (C_{in}, T, V), the ST-GCN blocks return a vector of shape $(C_{out}, T/8, V)$. $C_{in} = 2$ is the number of channels of each graph node, $T = 64$ is the length of the input sequences, $V = 15$ is the number of graph nodes (see the network structure described in Table II). $C_{out} = 256$ is the output channel. Thus the number of graph nodes V is reserved. Then the saliency vector $S \in \mathbb{R}^V$ is calculated as follows:

$$S = \sum_{t}^{T/8} \sqrt{\sum_{i}^{C_{out}} \left(O_{it}^2\right)} \tag{6}$$

where O_{it} is the elements of the output matrix O. The saliency vector shows the response of each graph node after spatial and temporal convolution. S_i can be considered as the weight index of Joint J_i (index i see Fig 2(a)). Considering the cooperation between joints during walking, the joints are divided into six groups: neck (J_0, J_1), torso (J_1, J_8), left arm (J_5, J_6, J_7), right arm (J_2, J_3, J_4), left leg (J_{12}, J_{13}, J_{14}) and right leg (J_9, J_{10}, J_{11}). In order to determine which body parts contribute more to the gait score prediction task and facilitate further study, the average saliency values of every group are calculated and ranked.

Based on the six groups proposed, we additionally choose 7 supplementary features to verify the effectiveness of the saliency value:

1) Neck Forward Bending Angle: the neck is represented by the link between Joint 0 and Joint 1, then the neck forward bending angle is obtained by calculating the angle between the neck and the vertical.

2) Torso Velocity: it is denoted by the horizontal velocity of Joint 1.

3) Torso Forward Bending Angle: the link between J_1 and J_8 represents the torso, then the torso forward bending angle is denoted by the angle between the torso and the vertical.

4) Arm Swing Amplitude: the arm swing amplitude is represented by the angle between the upper arm and the vertical.

5) Arm Swing Asymmetry (ASA): given the average swing amplitude of the left arm θ_L and the right arm θ_R, then $\theta_{\max} = \max(\theta_L, \theta_R)$, $\theta_{\min} = \min(\theta_L, \theta_R)$. The ASA is calculated as Equation (7).

$$ASA = \frac{45^\circ - \arctan(\theta_{\max}/\theta_{\min})}{90^\circ} \tag{7}$$

6) Step Length: it is represented by the max distance between two heels in a gait cycle.

7) Step Time: it is defined as the time interval between peaks and troughs of the heels distance-time curve (see Figure 2 (c)).

The Spearman correlation coefficients between the features and the gait scores are utilized to measure the contribution of each feature to the gait score.

III. Experimental Evaluation

A. Implementation Details

The cost function of the proposed model is the cross-entropy loss. By performing the grid search using the stochastic gradient descent (SGD) method, the optimal hyperparameters of the model are obtained. All training processes are initialized with a momentum of 0.9 and a learning rate of 0.01. All models are developed on the Pytorch framework and run on an Nvidia GeForce GTX 1660 Ti GPU with 6 GB memory. Each training epoch takes about 6 seconds, and the best performance is achieved at 280 epochs with a batch size of 32.

All experiments are conducted on an individual basis, which means the sub-clips of one subject cannot be separated by the train and evaluation splits. During the model validation, a voting mechanism is introduced for the subject-based evaluation [27]. The predicted score of one participant is determined by the majority of votes among its sub-clips. If the vote numbers of two scores are the same, the larger one is given.

B. Evaluation Metrics

All evaluations are conducted in a 5-fold cross-validation scheme. The per-class precision (Pre), recall (Rec), F1 score, and area under the ROC curve (AUC) are used to evaluate the proposed method. The confusion matrix and the accuracy (Acc) are used to indicate the total prediction results. In addition, as explained in [27], Cohen's κ coefficient is considered a more robust measure compared to the simple percent agreement measure since it considers the possibility that agreement occurs by chance. Hence we use the κ coefficient for comparison and evaluation. For all statistical analyses, $p < 0.05$ is considered a signal for statistical significance.

TABLE III
Classification Results of The Gait Score Predicting Task

	Prec	Rec	F1	AUC
Score-0	0.840	0.808	0.824	0.892
Score-1	0.586	0.654	0.618	0.745
Score-2	0.731	0.679	0.704	0.733
Total Acc			0.713	

TABLE IV
Ablation Study Results of The Proposed Method

	Prec	Rec	F1	AUC	κ
Proposed method	**0.713**	**0.710**	**0.710**	**0.789**	**0.569**
w/o skeleton stream	0.600	0.640	0.616	0.759	0.400
w/o silhouette stream	0.663	0.683	0.670	0.716	0.493
w/o LT-GEI	0.663	0.700	0.676	0.765	0.493
w/o proximal-distal	0.575	0.627	0.592	0.765	0.365

The use of **bold** font indicates the highest value among all comparisons for the same index. Unless explicitly stated, the convention of bold font is maintained throughout the rest of the paper.

C. Age Effects

ANOVA (Analysis of Variance) test and t-test are conducted to exclude the age effect on the subject groups (see Table I). Shapiro-Wilk test is done on each group to validate the normality of the age distributions. The p-values for the three groups are 0.66, 0.54, and 0.13, respectively, proving that the ages of all groups obey normal distribution. The ANOVA test on all three groups returned a p-value of 0.13, which indicates that there are no significant age differences across groups. The t-test between classes (0-1, 1-2, and 0-2) returned p-values of 0.288, 0.06, and 0.314, respectively, which suggest that no significant age difference is found between any two groups. According to the above results, the age effect is excluded from this study.

D. Classification Results and Ablation Study

In order to evaluate the method comprehensively, the sample-based results of each class are summarized in Table III and Figure 3. The overall classification accuracy is 71.25% with all AUC values larger than 0.7. Among the 54 patients in the dataset, only 4 are misclassified as healthy people, which means the sensitivity of the model to screen patients with PD has reached a sensitivity of 92.3%. In general, the results imply that the proposed model has achieved a result comparable to advanced relevant studies [24], [30].

Furthermore, to verify the necessity and effectiveness of all strategies proposed in Section II, ablation experiments are also conducted. The results are given in Table IV. The two-stream structure is verified by separately disabling the

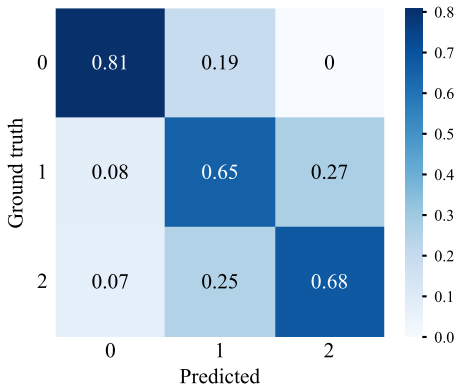


Fig. 3. Confusion matrix of the gait score prediction task.

TABLE V
COMPARISON WITH BASELINE METHODS

	Prec	Rec	F1	AUC	κ
Our method	**0.713**	0.710	**0.710**	**0.789**	**0.569**
SVM	0.538	0.598	0.555	0.676	0.308
Random Forest	0.587	0.632	0.602	0.760	0.382
2s-AGCN ([43])	0.625	**0.732**	0.661	0.702	0.437
3DCNN ([24])	0.588	0.588	0.588	0.721	0.382
LSTM ([44])	0.613	0.638	0.621	0.773	0.419

skeleton and silhouette streams. The precision decreases significantly (21.3%) when the skeleton stream is disabled, while the AUC value drops greatly (7.3%) when the silhouette stream is disabled. It seems that the skeleton stream works extremely well in dominant categories, and the silhouette stream has a more balanced performance for all categories. The fusion of these two networks is proven to be more efficient compared to each of them. Then the LT-GEI is replaced by GEI to put into the silhouette stream. All aforementioned evaluation indicators decrease, especially the κ coefficient. It implies that the addition of the LT-GEI increases the robustness of the model. Finally, we substitute the proximal-distal configuration with the uniform configuration (all graph edges are considered to be the same). It is worth noting that the model without the proximal-distal configuration works even worse than the skeleton stream itself with a heavy margin, with the exception of the AUC value.

E. Comparison With Baseline Methods

To further demonstrate the performance of the proposed model, several baseline methods are examined on the same clinical dataset (see Table V), including Support Vector Machine (SVM), Random Forest (RF), the SOTA skeleton-based action recognition model 2s-AGCN [43], LSTM [44] and the silhouette-based Parkinsonian gait classification model 3DCNN [24]. The F-1 score, AUC, and κ are three comprehensive evaluation metrics. For the F-1 score, the proposed method outperformed the other methods by 15.5%, 10.8%, 4.9%, 8.9%, and 12.2%. For the AUC, the proposed method outperformed the other methods by 11.3%, 2.9%, 8.7%, 1.6%, and 6.8% (see Table V). The κ coefficient of the proposed model is 0.569, which is significantly higher than the other four compared methods.

TABLE VI
SUPPLEMENTARY FEATURES

Location	Rank	Metric	Feature	Spearman Corr.
Neck	3	Forward angle	Average	**0.43**
Torso	6	Forward angle	Average	0.16
	2	Velocity	Average	**0.74**
Arm	1	Swing amplitude	Maximum	**0.78**
	4	Swing asymmetry	ASA	0.27
Leg	7	Step length	CV	0.04
	5	Stride time	CV	0.26

The average represents the average value of the features in all frames.
The maximum represents the maximum value of the arm swing amplitude of both sides in all frames.
The CV represents the coefficient of variability of the leg related features in all gait cycles.
The top three Spearman correlation coefficients are bolded.

F. Saliency Analysis

The average saliency values among all PD patients are shown in Figure 4 (a). It can be observed that the upper body shows a higher response to the ST-GCN blocks. The neck (Joint 0), head (Joint 1) and left shoulder (Joint 2) contribute the most to the prediction of gait scores. The average saliency values of the proposed six groups are given in Figure 4 (b). It is obvious that the neck, torso and left arm are the three body parts with the largest saliency values.

As for the supplementary features, the three most relevant features are the arm swing amplitude, the torso velocity, and the neck forward bending angle, with Spearman correlation coefficients of 0.78, 0.73, and 0.43, respectively. Moreover, among all proposed supplementary features, only these three features are found to be significantly correlated to the gait score ($p < 0.05$). The statistical results show that the body parts concerned more by the model are indeed more important for the prediction task. Thus the efficiency of the proposed saliency value is proved.

IV. DISCUSSION

The main objective of this study is to develop a video-based Parkinsonian gait assessment system in two principal ways: (1) develop a skeleton-silhouette fusion convolutional network to predict the patients' UPDRS gait scores and (2) extract network-derived supplementary features to provide credible quantification metrics with higher resolution. This section will be discussed in terms of these two objectives and how the experimental results compare to those of prior related studies.

A. Classification of Parkinsonian Gait

Accurate assessment of gait impairments for patients with mild PD symptoms (i.e., with a UPDRS gait score of 1 or 2) is important for efficient treatment and rehabilitation. Meanwhile, studies have shown that it is more difficult to accurately classify UPDRS gait scores of 1 and 2 compared to other

TABLE VII
COMPARISON WITH SOTA VIDEO-BASED METHODS

Studies	Subjects	Video Clips (0/1/2)	Model Input	Precision	Recall	Specificity	F-1 Score
Present Study	80	156/156/168	2D skeleton, silhouette	**0.713**	**0.710**	**0.853**	**0.710**
Guo *et al.*[26]*	**142**	38/194/154	2D skeleton	0.582	0.553	-	0.562
Mandy *et al.*[27]*	51	-	3D skeleton	0.667	0.683	0.787	0.661
Sabo *et al.*[30]	53	88/131/180	3D skeleton	0.529	0.553	0.670	0.520

In the "Video Clips" column, the numbers of gait videos with UPDRS gait scores of 0, 1, and 2 are listed in the table.
*In studies [26] and [26], patients with a gait score of 3 are also included, while not in the other studies. In this table, the UPDRS-0,1,2 classification results presented in [26] and [27] are directly derived (not retraining the models because of lack of related datasets) from their experimental data to ensure the same benchmark for comparison.

scores [26], [27]. There is a limited number of video-based methods for predicting MDS-UPDRS gait scores which can provide a relatively high prediction accuracy. To show the performance of the proposed approach for predicting the gait scores of 1 and 2, a comparison with some latest video-based studies on Parkinsonian gaits, such as [26], [27], and [30], is carried out (see TableVII). It should be noted that patients with UPDRS gait score 3 are also included in [26] and [27], while in [30] and the present study, they are not recruited. To compare under a same benchmark, only the UPDRS-0,1,2 classification results are calculated according to the experimental data provided by [26] and [27].

Guo et al. [26] proposed a neural network named 2s-ST-AGCN for the assessment of Parkinsonian gait. 441 gait examination video clips of 142 patients with gait scores of 0, 1, 2 and 3 were collected and tested. Based on the 2D poses estimated by the AlphaPose [45], the method achieved a balanced recall of 65.66% in the 4-class dataset. The F-1 score of class-0, 1 and 2 were 72.3%, 57.3%, and 50.8% respectively, thus the balanced F-1 score for classes 0-2 was 56.2%.

Lu et al. [27] proposed the OF-DDNet (ordinal focal double-feature double-motion network) to handle class imbalance and noisy labels of the constructed dataset. In the MDS-UPDRS gait score estimation task, the number of subjects with scores 0, 1 2 and 3 were 10, 33, 8 and 4, respectively. 3D skeleton sequences were extracted from gait videos with VIBE (Video Inference for human Body pose and shape Estimation) [46]. Benefit from the focal loss, the class imbalance was alleviated, and a balanced accuracy of 72% was achieved on the 4-class dataset. The classification performance of each class can be derived from the confusion matrix given in the study, a balanced F-1 score of 66.1% was achieved for classes 0-2.

Sabo et al. [30] proposed a two-stage training approach to evaluate the Parkinsonism of individuals with dementia. 5 types of human pose (i.e. OpenPose, Detectron, AlphaPose, Microsoft Kinect 2D & 3D) and 2 types of models (i.e. OF-DDNet and ST-GCN) were tested on the 53 subjects dataset. Note that only patients with UPDRS gait scores of 1 and 2 were included in the study. The best performance was achieved using Kinect 3D pose and the ST-GCN, with a balanced F-1 score of 52%.

Class imbalance in the clinical dataset is commonly seen in previous studies. In [27], the focal loss was leveraged to combat the problem. In the present study, we construct a class-balanced dataset using patients with gait scores of 1 and 2. In addition, some gait impairments (e.g., reduced strike speed) show a significant link with age. Eliminating the age effects is crucial to ensure an accurate evaluation of the UPDRS classification methods. However, it is rarely taken into account in prior studies. In the present study, there are no significant age differences across all three groups of the dataset, which indicates that our results are not related to age. Finally, the classification results exceed the best performance of the previous studies [26], [27], [30] on 4 comparative metrics including precision (71.3% vs. 66.7%), recall (71% vs. 68.3%), specificity (85.3% vs. 78.7%) and F-1 score (71% vs. 66.1%). We have also tested the proposed model on a balanced dataset with 26 healthy subjects and 26 PD patients, the total accuracy was 90.4%.

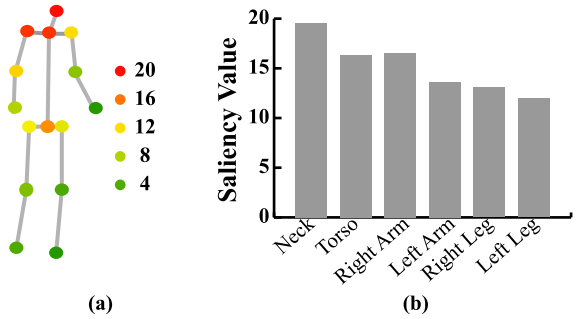


Fig. 4. The saliency values of the joints and body parts.

B. Saliency Values and Supplementary Features

Prior studies have analyzed some key gait features for PD assessment, such as neck bending [40], gait speed, stride length [47], arm swing amplitude and asymmetry [9], using motion capture, inertial sensor and accelerometer. Most of them focused on analyzing the differences between healthy individuals and PD patients. In the present study, seven video-based supplementary features are extracted and evaluated, while three of them are found to be strongly correlated with the gait scores, i.e., the arm swing amplitude, the torso velocity and the neck forward bending angle (see Section III-F). The results show great promise in developing quantitative means with higher resolution for more precise parkinsonian gait assessments.

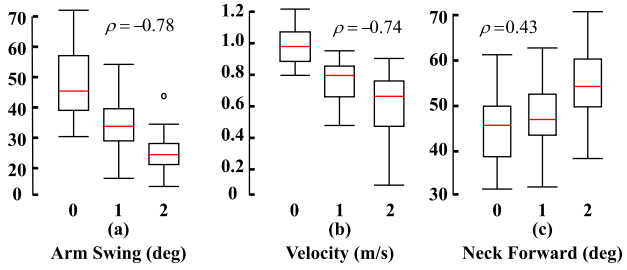


Fig. 5. Boxplots of two proposed supplementary features with the gait scores. (a) Max arm swing amplitude. (The black circle represents the outliers.) (b) Gait velocity. (c) Neck forward bending. The ρ represents the Spearman correlation coefficient between the features and the gait scores.

The arm swing amplitude is found to be different between PD patients and healthy controls in [47]. The typical cases of participants with scores of 0,1,2 are shown in Figure 6(a)-(b). It can be seen that compared to PD patients, the arm swing amplitudes of healthy subjects are larger, and the cadence and bilateral symmetry are also better. However, the arm swing asymmetry feature shows a low correlation with the gait scores. One reason is that most of the patients are diagnosed as H&Y stage 2 and above. Only two patients are diagnosed as H&Y stage 1. According to [5], H&Y stage 2 is defined as bilateral involvement without impairment of balance, which means most of the patients involved are affected by PD on both sides. Thus it is hard to determine the relationship between the bilateral arm swing and the gait scores. On the other side, there is a significant correlation between the maximum value of the arm swing amplitude and the gait UPDRS score. The max arm swing amplitude boxplot of three groups is provided in Figure 5(a). It can be seen that the median of each group (red lines) shows a decreasing trend with the gait UPDRS scores. And the Spearman correlation coefficient is $\rho = -0.78$, suggesting a significant negative correlation.

Many studies have shown that PD patients walk slower than healthy people [38], [47]. In the present study, the horizontal velocity of the torso is chosen to be a reflection of gait speed, and shows a significant negative correlation with the gait score (Spearman correlation coefficient $\rho = -0.74$; see Figure 5 (b) and Fig 6 (c)) Note that there exists persistent oscillation in the velocity curve in Fig 6 (c), since the differential operation and the short sampling period (0.04s) amplified the noise introduced by skeleton estimation. However, the noise can be compressed by averaging. Compared to prior sensor-based methods detecting gait speed by measuring foot motions, the proposed method can avoid the common error accumulation problem.

In [48], a summary index composed of neck and trunk forward bending angle is proposed to evaluate postural abnormalities severity in Parkinson's Disease. A significant correlation was detected between the index and the total UPDRS scores (Spearman correlation coefficient $\rho = 0.37$). In the present study, the relationship between the neck forward bending angle and the UPDRS gait score is evaluated. The

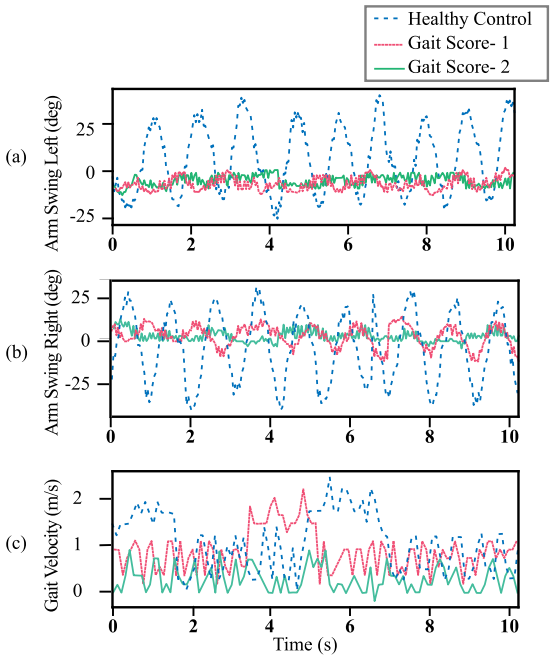


Fig. 6. Plots of the supplementary features extracted from the gait videos. (a) Left arm swing angles. (b) Right arm swing angles. (c) Gait velocity.

Spearman correlation coefficient between the neck forward bending angle and the gait score is 0.43, and the bending angle shows a positive correlation with the gait score (see Figure 5 (c)).

C. Clinical Application and Future Work

Existing Parkinson's assessment methods have variously been proposed as substitutes for clinical testing. The proposed method can not only predict the MDS-UPDRS gait scores but also provide fine-grained supplementary features for high-resolution gait quantification. The earliest stages of Parkinson's disease can be difficult to recognize, and assessments administered in clinics can be episodic because many people with Parkinson's disease may respond differently at home than in the hospital. Moreover, the gaits of PD patients are influenced by factors such as medication intake and DBS treatments [49]. Benefit from the smartphone-based data collection, the system holds significant future potential in the home-based diagnosis of early PD, long-term home-based assessment of the patient's symptoms and response to medical intervention as well as closed-loop feedback for DBS stimulation adjustment, which are of great importance in the clinical management of PD. In future works, more PD patients will be recruited to test the generalization capability and robustness of the proposed method. In the present study, patients with Scores 3 and 4 are excluded. In the future, patients in these classes will be involved. Furthermore, the measurement accuracy of the supplementary features will be tested in detail.

V. Conclusion

In this study, a video-based method for automatic and quantitative assessment of gait impairments in PD

using a skeleton-silhouette fusion neural network. The skeleton-silhouette neural network is adopted to extract features from the gait videos and predict the UPDRS gait scores. To understand which body parts contribute more in correctly predicting the gait scores, the saliency values are derived from the neural network and ranked. Finally, several supplementary features are designed and examined for the verification of the saliency values. Extensive experiments are done on a clinical dataset with 80 participants. The results show that satisfactory prediction accuracy can be achieved using the proposed method. In addition, three supplementary features are found to be significantly correlated to the gait scores, thus can be used as gait disorder quantification metrics with higher resolution. The data collection can be easily done through smartphones, and nearly no clinical training is required. In future work, the proposed method will be tested for home-based early PD diagnosis. More patients will be involved to examine the generalization capability and robustness of the proposed system. Meanwhile, the measurement accuracy of the supplementary features will be tested to guarantee the soundness and effectiveness of the quantification metrics.

ACKNOWLEDGMENT

The authors would like to thank all subjects and clinicians at Tianjin Huanhu Hospital for their kind cooperation in this study, as well as Huawei MindSpore and Tianjin AI Computer Center for their support in training an online model based on the proposed method.

REFERENCES

[1] E. R. Dorsey et al., "Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the global burden of disease study 2016," *Lancet Neurol.*, vol. 17, no. 11, pp. 939–953, Nov. 2018.

[2] R. Z. U. Rehman, S. Del Din, Y. Guan, A. J. Yarnall, J. Q. Shi, and L. Rochester, "Selecting clinically relevant gait characteristics for classification of early Parkinson's disease: A comprehensive machine learning approach," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Nov. 2019.

[3] R. L. Nussbaum and C. E. Ellis, "Alzheimer's disease and Parkinson's disease," *New England J. Med.*, vol. 348, no. 14, pp. 1356–1364, Apr. 2003.

[4] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh, "Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging," *Exp. Syst. Appl.*, vol. 41, no. 7, pp. 3333–3342, Jun. 2014.

[5] C. G. Goetz et al., "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, Nov. 2008.

[6] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 50, no. 2, p. 318, Feb. 1998.

[7] Y. Xia, Z. Yao, Q. Ye, and N. Cheng, "A dual-modal attention-enhanced deep learning network for quantification of Parkinson's disease characteristics," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 42–51, Jan. 2020.

[8] X. Peng et al., "Gait analysis by causal decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 953–964, 2021.

[9] M. D. Lewek, R. Poole, J. Johnson, O. Halawa, and X. Huang, "Arm swing magnitude and asymmetry during gait in the early stages of Parkinson's disease," *Gait Posture*, vol. 31, no. 2, pp. 256–260, Feb. 2010.

[10] A. Plate, D. Sedunko, O. Pelykh, C. Schlick, J. R. Ilmberger, and K. Bötzel, "Normative data for arm swing asymmetry: How (A) symmetrical are we?" *Gait Posture*, vol. 41, no. 1, pp. 13–18, Jan. 2015.

[11] H. Wang, M. E. Esi Acquah, X. Zhang, Q. Xu, W. Chen, and D. Gu, "The effect of visual cues on muscular activation in the lower limbs of Parkinson's disease patients with freezing of gait: A preliminary study," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 6211–6214.

[12] F. Cheng, Y. Yang, Y. Wu, C. Lu, S. Cheng, and R. Wang, "Beta event-related desynchronization can be enhanced by different training programs and is correlated with improved postural control in individuals with Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 10, pp. 1957–1964, Oct. 2018.

[13] P. B. Shull, W. Jirattigalachote, M. A. Hunt, M. R. Cutkosky, and S. L. Delp, "Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention," *Gait Posture*, vol. 40, no. 1, pp. 11–19, May 2014.

[14] L. Lonini et al., "Wearable sensors for Parkinson's disease: Which data are worth collecting for training symptom detection models," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–8, Nov. 2018.

[15] E. M. J. van Brummelen et al., "Quantification of tremor using consumer product accelerometry is feasible in patients with essential tremor and Parkinson's disease: A comparative study," *J. Clin. Movement Disorders*, vol. 7, no. 1, pp. 1–11, Dec. 2020.

[16] L. Di Biase et al., "Quantitative analysis of bradykinesia and rigidity in Parkinson's disease," *Frontiers Neurol.*, vol. 9, p. 121, Mar. 2018.

[17] R. A. Ramdhani, A. Khojandi, O. Shylo, and B. H. Kopell, "Optimizing clinical assessments in Parkinson's disease through the use of wearable sensors and data driven modeling," *Frontiers Comput. Neurosci.*, vol. 12, p. 72, Sep. 2018.

[18] W. Huo et al., "A heterogeneous sensing suite for multisymptom quantification of Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1397–1406, Jun. 2020.

[19] L. Formstone, W. Huo, S. Wilson, A. McGregor, P. Bentley, and R. Vaidyanathan, "Quantification of motor function post-stroke using novel combination of wearable inertial and mechanomyographic sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1158–1167, 2021.

[20] L. Dranca et al., "Using Kinect to classify Parkinson's disease stages related to severity of gait impairment," *BMC Bioinf.*, vol. 19, no. 1, pp. 1–15, Dec. 2018.

[21] Z. Zhang et al., "Automated and accurate assessment for postural abnormalities in patients with Parkinson's disease based on Kinect and machine learning," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, pp. 1–10, Dec. 2021.

[22] J. E. Pompeu et al., "Feasibility, safety and outcomes of playing Kinect adventures!Ů for people with Parkinson's disease: A pilot study," *Physiotherapy*, vol. 100, no. 2, pp. 162–168, Jun. 2014.

[23] G. Palacios-Navarro, I. García-Magariño, and P. Ramos-Lorente, "A Kinect-based system for lower limb rehabilitation in Parkinson's disease patients: A pilot study," *J. Med. Syst.*, vol. 39, no. 9, pp. 1–10, Sep. 2015.

[24] L. C. Guayacán and F. Martínez, "Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network," *J. Biomed. Informat.*, vol. 123, Nov. 2021, Art. no. 103935.

[25] Y. Yang et al., "A video-based method for assessment of hip-knee-ankle coordination during walking," in *IEEE 11st Annu. Int. Conf. Cyber. Technol. Autom. Control Intell. Syst. (CYBER)*, Jul. 2021, pp. 608–613.

[26] R. Guo, X. Shao, C. Zhang, and X. Qian, "Multi-scale sparse graph convolutional network for the assessment of Parkinsonian gait," *IEEE Trans. Multimedia*, vol. 24, pp. 1583–1594, 2022.

[27] M. Lu et al., "Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos," *Med. Image Anal.*, vol. 73, Oct. 2021, Art. no. 102179.

[28] R. Guo, X. Shao, C. Zhang, and X. Qian, "Sparse adaptive graph convolutional network for leg agility assessment in Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2837–2848, Dec. 2020.

[29] T. T. Verlekar, H. De Vroey, K. Claeys, H. Hallez, L. D. Soares, and P. L. Correia, "Estimation and validation of temporal gait features using a markerless 2D video system," *Comput. Methods Programs Biomed.*, vol. 175, pp. 45–51, Jul. 2019.

[30] A. Sabo, S. Mehdizadeh, A. Iaboni, and B. Taati, "Estimating parkinsonism severity in natural gait videos of older adults with dementia," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2288–2298, May 2022.

[31] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[32] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.

[33] T. T. Verlekar, P. L. Correia, and L. D. Soares, "View-invariant gait recognition system using a gait energy image decomposition method," *IET Biometrics*, vol. 6, no. 4, pp. 299–306, Jul. 2017.

[34] M. Yamamoto, K. Shimatani, M. Hasegawa, Y. Kurita, Y. Ishige, and H. Takemura, "Accuracy of temporo-spatial and lower limb joint kinematics parameters using OpenPose for various gait patterns with orthosis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2666–2675, 2021.

[35] C. A. Putnam, "A segment interaction analysis of proximal-to-distal sequential segment motion patterns," *Med. Sci. Sports. Exerc.*, vol. 23, no. 1, pp. 130–144, 1991.

[36] X. Cao et al., "Video based shuffling step detection for parkinsonian patients using 3D convolution," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 641–649, 2021.

[37] J. C. Pérez-Ibarra, A. A. G. Siqueira, and H. I. Krebs, "Identification of gait events in healthy subjects and with Parkinson's disease using inertial sensors: An adaptive unsupervised learning approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2933–2943, Dec. 2020.

[38] L. Alcock, B. Galna, R. Perkins, S. Lord, and L. Rochester, "Step length determines minimum toe clearance in older adults and people with Parkinson's disease," *J. Biomechanics*, vol. 71, pp. 30–36, Apr. 2018.

[39] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. 32nd AAAI Conf. Artif. Intell.*, Apr. 2018, pp. 7444–7452.

[40] M. Tinazzi et al., "Validity of the wall goniometer as a screening tool to detect postural abnormalities in Parkinson's disease," *Parkinsonism Rel. Disorders*, vol. 69, pp. 159–165, Dec. 2019.

[41] M. Welling and T. N. Kipf, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1–14.

[42] R. Sharma, L. Pillai, A. Glover, and T. Virmani, "Objective impairment of tandem gait in Parkinson's disease patients increases with disease severity," *Parkinsonism Relat. Disord.*, vol. 68, pp. 33–39, Nov. 2019.

[43] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2019, pp. 12026–12035.

[44] A. Shahroudy, J. Liu, T. Ng, and G. Wang, "NTU RGB+D: A large scale dataset for 3D human activity analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1010–1019.

[45] H. Fang, S. Xie, Y. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2353–2362.

[46] M. Kocabas, N. Athanasiou, and M. J. Black, "VIBE: Video inference for human body pose and shape estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5252–5262.

[47] A. Sánchez-Rodríguez et al., "Sensor-based gait analysis in the premotor stage of LRRK2 G2019S-associated Parkinson's disease," *Parkinsonism Rel. Disorders*, vol. 98, pp. 21–26, May 2022.

[48] R. Hong et al., "A summary index derived from Kinect to evaluate postural abnormalities severity in Parkinson's disease patients," *NPJ Parkinson's Disease*, vol. 8, no. 1, pp. 1–8, Aug. 2022.

[49] B. R. Bloem, M. S. Okun, and C. Klein, "Parkinson's disease," *Lancet*, vol. 397, no. 10291, pp. 2284–2303, 2021.